הפרויקט : אפיון והצעת חלופות בנוגע לפערים במערכת החינוך בהישגים ובזכאות לבגרות

חברי קבוצה:
א. אמרי דרור, imri.dror, imri.dror@mail.huji.ac.il
ב. מעיין שרון, Maayan.sharon , Maayan.Sharon@mail.huji.ac.il

1. תיאור הבעיה:

מערכת החינוך מהווה נדבך מרכזי בעיצוב העתיד של צעירי ישראל, ולהצלחה במערכת זו השפעה ניכרת על חייהם הבוגרים (קבלה לתפקידים משמעותיים בצבא, השכלה גבוה, כושר השתכרות..). על אף שהמערכת מקבלת תקצוב לכל תלמיד, קיימים פערים בתקצוב ובשירותי החינוך באזורים שונים בארץ. אנו מעריכים כי ניתן למצוא דפוסים במידע שיאפשרו איתור פערי מדיניות או אתגרים שיש להתמקד בהם - ע"ב השוואה בין פרטי מידע שונים על תקציב, שעות לימוד, ומאפייני המסגרת והדמוגרפיה של בתי הספר (ממלכתי / דתי / חרדי, אזור מגורים ונפה, חתך סוציואקונומי, מגזר וכו').

2. תיאור של הדאטה:

א. 13 קבצי אקסל הכוללים חיתוכים שונים מהאתר של משרד החינוך (היפר-קישור):
גודל הקבצים - סה"כ MB16.2, טבלאות אשר נגזרו ככל הנראה ממקור מרכזי אחד (או שניים לכל היותר), המכילים נתוני תפעול בתי הספר (תקציב, שעות חינוך, היקפי כיתות וכו') ושל תוצאות במדדים חיצוניים (מבחני מיצ"ב, תוצאות בגרות, אחוזי גיוס לצבא ולשירות לאומי). לאחר ניתוח הנתונים התמקדנו בטבלאות זכאות לבגרות ותקציב בתי הספר.

ב. קובץ אקסל אחד של נתוני זכאות לבגרות 2013-2016 (היפר-קישור):
גודל הקובץ - KB8860, טבלה בת 23896 שורות, כל שורה מייצגת נתונים של בית ספר באחת השנים הנ"ל. בקובץ 53 עמודות המייצגות נתונים שונים על בית הספר.

ג. נתונים רשמיים ממשרד החינוך על מיקומי בתי ספר: קובץ אקסל המכיל מידע רשמי של משרד החינוך על בתי הספר, אליו נוספו מיקומים גיאוגרפים לפי כתובת של כל בית ספר. בגודל KB356. בקובץ 21 עמודות המייצגות נתונים שונים של בתי הספר, על כ2089 שורות (כל שורה- בית ספר). את הקובץ קיבלנו באדיבותו של חברנו לספסל הלימודים - עציון הררי.

3. הפתרון שלנו:

אנו התמקדנו בפתרונות של ניתוח ואגרגציה של הנתונים, והנגשה שלהם כתוצרים ניתנים להבנה ולשימוש. מטרתנו לבדוק את תוצאות מערכת החינוך בזכאות לבגרות למול מאפייני תקציב, פיקוח ושירותי חינוך נוספים לבתי ספר ממגזרים ומעמד סוציואקונומי שונים. זאת במטרה לאפשר ערעור על פרדיגמות קיימות ויצירת עזרים לגיבוש מדיניות חדשה מבוססת נתונים.
בהתאם, התוצר שלנו מורכב משני חלקים:

1. ניתוח וסקירת מאקרו מעמיקה על התפלגויות בתי הספר לפי שנות הקמה, חלוקה למאפיינים דמוגרפים שונים, ניתוח היבטים תקציביים שלהם, וניתוח כללי של זכאויות לבגרות והפערים ביניהם. זאת במטרה הן להבין את המידע והן כתיאור כללי של מערכת החינוך "ממעוף הציפור".
2. שימוש במספר מודלים לתיאור וניבוי זכאות לבגרות לפי יתר מאפייני בתי הספר, והשוואתם לחלוקות הקיימות של בתי הספר ל"קבוצות דומים" - בכך:
    1. חלוקה לתתי קבוצות לפי שילובי מאפיינים חדש (מגזר ואחוזון טיפוח).
    2. קלאסטרים המבוססים על כלל נתוני הזכאות לבגרות.
    3. עץ החלטה לניתוח התפלגות בתי הספר לפי מאפיין נמדד.

זאת תוך הנגשה של התוצרים כמידע טבלאי, תוצרים ויזואלים וממשקים אינטראקטיביים - כשכבות על מפה, וכממשק ליצירת עץ החלטות.
התבססנו על המודלים הקיימים בספריות "pandas", "scikit learn" , "geopy", "numpy", ו-"matplotlib" ושירותים של GOOGLE MAPS תוך התאמה שלהם לצרכינו:

- מודל K-Means בגדלים המתאימים למידע (ע"ב ניתוח SSE וSilhouette)
- מודל עץ החלטות - תוך הגבלת עומק, לאור הצורך בקריאות ואפשרות של היעדר הגעה לסיום עם סינגלטונים.
- ניתוח מילים ליצירת גרפים ועננים מילים (בעזרת אתר https://www.wordclouds.com/)
- שימוש ב MY MAPS, שירות של GOOGLE MAPS על מנת להעלות את הנתונים המעובדים עבור ערים ובתי ספר על מפה.

בהתאם לאמור, יצרנו מודלים שונים של הערכת אמינות התוצאות:

1. בחינת נתוני הניתוח למול נתונים גולמים בחלוקה לערים ורשויות.
2. שילוב של דגימה והשוואה ידנית של מקרי בוחן מהתוצרים לבחינת מציאות.
3. בחנו את הנחות המוצא שלנו למול גורמים מהתחום (קרובי משפחה העובדים משרד החינוך).
4. כמו כן השתמשנו בהשוואות שונות בין המודלים (עליהן נרחיב בפרק התוצרים על כל מודל) כדי לבדוק את תוצאות הניתוחים השונים והקוהרנטיות ביניהם.

פרק תוצרים

תוצרי ניתוח וסקירת מאקרו על בתי הספר:

*יש לציין כי בכל קבצעי הקוד נעזרנו בספריית pandas לניתוח וספריית matplotlib ליצירת גרפים

* לכל סעיף בתוצרים מצורפת תקייה עם התוצרים עצמם.

1. **שמות בתי ספר כמאפיין בסיסי לניתוח:**

a. **קובץ קוד:** schoolNames.py

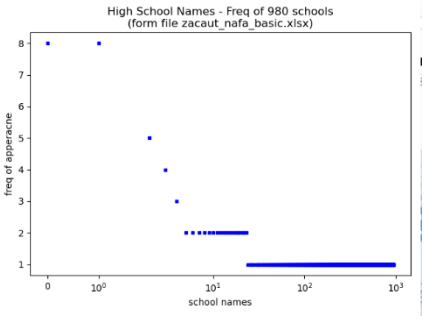
b. **מטרה:** אפיון כללי של בתי הספר בארץ לפי שמם – המהווה את פריט הידע הבסיסי ביותר הנתון לאדם המתעניין בבית הספר.

c. **הסבר התהליך:** יצרנו רשימה עם שמות כל בתי הספר, הורדנו סימני פיסוק פירקנו את השמות למילים והתייחסנו לכל מילה כ'טוקן' בכוונה לאתר תבניות הקשורות לשמות בתי הספר. לאחר מכן יצרנו מילונים עם המילה וכמות ההופעות שלה ולשם ויזואליזציה נעזרנו באתר https://www.wordclouds.com/

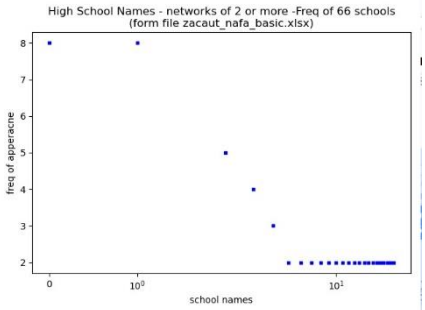
d. **בדיקת מהימנות:** היות והשמות המלאים אינם חד-חד ערכיים, אנו משתמשים בסמל מוסד להבחנה בין בתי הספר.

e. **תוצרים ומסקנות:**

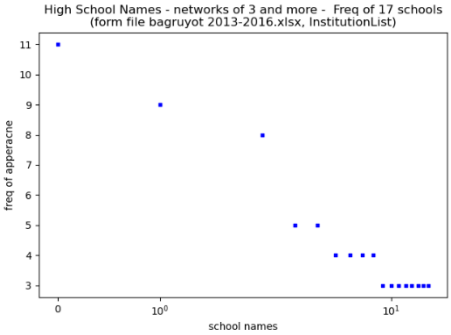
i. מבחינה של כלל שמות בתי הספר בולטים המון שמות ייחודיים, ובמקרים שאינם כאלו – מדובר ברשות או במילים פוסלות (חינוך, בי"ס וכו') או מאפיינות (לדוגמא ציבור דתי "בנע","דתי", "אמית".



ii. מהרשימה שקיבלנו בשקף קודם, הורדנו את כל בתי ספר עם שם יחיד כדי לקבל 'רשתות' (רשת: שני בתי ספר ומעלה בעלי אותו השם) ואיתרנו 66 בתי ספר המהווים כ24 רשתות. מעניין לראות ריבוי שמות עם מאפיינים דתיים (מבדיקה ידנית כעשרה מתוך 24 שמות יחודיים).



iii. לאור ריבוי השמות עם מאפיינים דתיים, בחינה דומה של קובץ המכיל גם בתי ספר שאינם זכאים לבגרות או שאין לנו עליהם מידע מלא מעלה תמונה דומה (אם כי יותר ברורה, כאשר הרוב המוחלט של הרשתות הן דתיות). במקרה זה צימצום לרשתות בעלי 3 בתי ספר ומעלה העלה יחס אפילו יותר ברור של 8 בתי ספר דתיים מתוך 17, מה שעשוי להעיד על בולטות של רשתות חינוך בסקטור הדתי והחרדי.



2. **ניתוח פתיחת בתי ספר חדשים**

a. **מטרה:** אפיון התפלגות מועדי פתיחת בתי הספר כמדד לקיבעון או צמיחה של החינוך בסוגי הפיקוח ובמגזרים השונים.

b. **הסבר התהליך:** ניתחנו את התפלגות הנתונים לאורך השנים ובחנו אותה בהתעמקות לפי מאפיינים.

c. **בדיקת מהימנות:** לאור ההעמקה בתתי קבוצות, בחנו את התוצאות למול ידע על ההיסטוריה של החברה הישראלית.

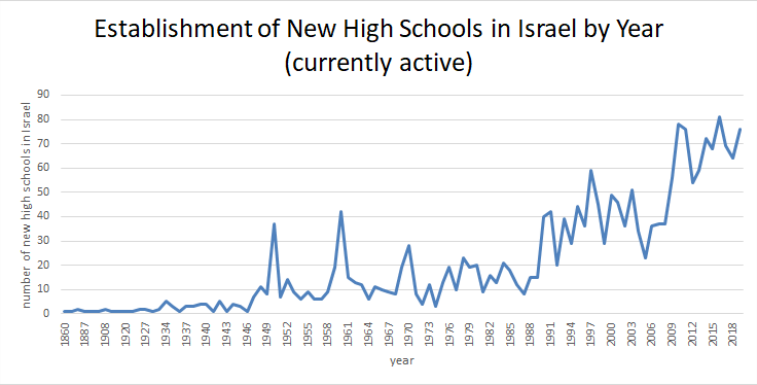
d. **תוצרים ומסקנות** (שליפת מידע ראשונה בקובץ newSchool.py. תוצר מלא ניתן לראות בקובץ new schools by years.xlsx):

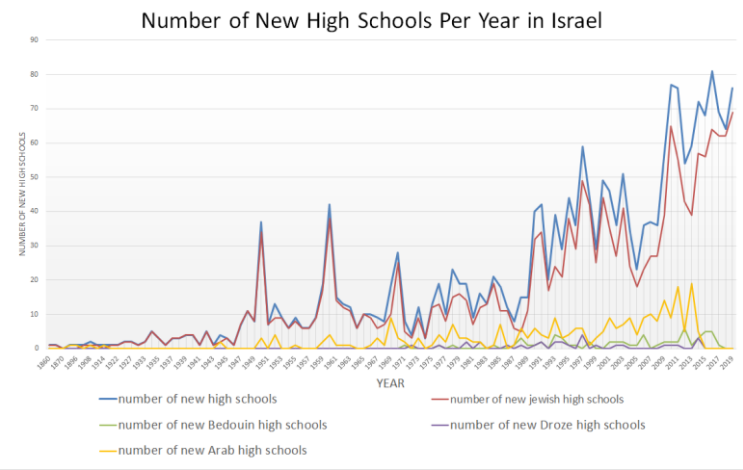
i. כלל בתי הספר לפי שנה – ניתן לראות עליה חדה בשנה הראשונה להקמת המדינה, או לדוגמא בשנים 1958-1964 עליה חדה (מתאים למספרים הרשמיים של העליות באותה תקופה, תוספת של כ300 אלף עולים בנוסף לגידול הטבעי), גרף ראשון.

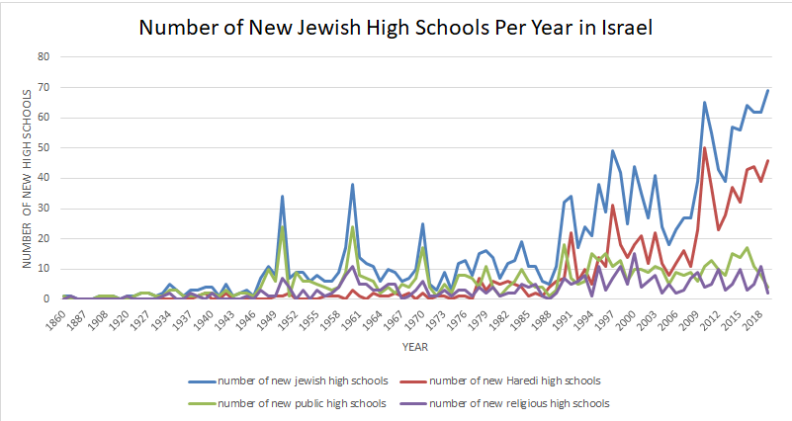
ii. פילוחים לפי שנה, שנה ומגזר – בהם ניתן לראות דומיננטיות ברורה של המגזר היהודי בהשפעה כמותנית על מגמות פתיחת בתי הספר לאורך השנים (בגרף השני: העקום הכחול – הכללי לעומת האדום – היהודי).

iii. מבחינה ממוקדת במגזר היהודי בחלוקה לפי סוגי פיקוח (ממלכתי / ממלכתי דתי / חרדי), גרף שלישי, ניתן לשים לב כי עד שנות השמונים, רוב בתי הספר שנפתחו השתייכו לזרם הממלכתי (כלומר אוכלוסייה חילונית, עקומה ירוקה) עם השפעות של הזרם הממלכתי-דתי, ומתחילת שנות התשעים ישנה מגמה ברורה של התחזקות הזרם החרדי, כאשר מספר בתי הספר הנפתחים לטובת ציבור זה גדול באופן משמעותי מכל זרם אחר במגזר היהודי. מעניין לראות כי במקביל מספר התיכונים החדשים הנפתחים לטובת הציבור החילוני נמצא במגמת עלייה איטית, ולעיתים אף בירידה (בהתחשב בגידול דמוגרפי טבעי).

iv. ישנם מספר הסברים לתופעה זאת. הראשון הוא גידול דמוגרפי מהיר של הזרם החרדי ביחס לאוכלוסייה הכללית, הדורש בהתאם יותר מוסדות לימוד. הסבר נוסף הוא המגמה של הממשלה בשני העשורים האחרונים לשלב לימודי ליבה בגילאי התיכון של הזרם החרדי. עניין זה דרש פתיחת בתי ספר חדשים המותאמים ללימודי קודש וחול, במתכונת מתאימה לציבור זה. בהתבוננות נוספת בנתונים שאספנו זיהינו כי מספר לא מבוטל של בתי ספר תיכונים המשתייכים לזרם החרדי מכילים מספר קטן של תלמידים (פחות מ-100) ביחס לבתי ספר המשתייכים לזרם החילוני, מה שכנראה משפיע על הגידול בכמות בתי הספר.







3. **זכאות לבגרות לפי ערים ונפות גאוגרפיות:**

a. **קובץ קוד:** zacaut2.py מורכב מסקריפטים שקיבצנו יחד לשם נוחות, בפועל כל חלק (מסומן בדוקומנטציה) רץ בנפרד.

b. **מטרה:** אפיון התפלגות הזכאות לבגרות בראי נתונים גיאוגרפיים, כמוקד משפיע פוטנציאלי, וכבסיס לתוצר כשכבת מפה שתאפשר אקספלורציה, גמישות למשתמש והשוואה מבוססת מיקום.

c. **הסבר התהליך:**

i. **עבור הנפות:** איחדנו בין מאגרי המידע על זכאות בתי ספר (school – bagrut.xlsx) ועל שייכות לנפה (schools_utm_only_hs.xlsx) וחישבנו ממוצע מסוכם לכל נפה, בכל שנה ובממוצע על כל השנים הנתונות. קובץ סופי - zacaut_per_nafa_mean.xlsx

ii. **עבור הערים:** יצרנו קובץ עזר - city_mean_zacaut_with_nafa.xlsx בו חישבנו את ממוצע זכאות לכל עיר וגם הצמדנו לכל עיר את הנפה המתאימה לחישובים עתידים. יצרנו קובץ עזר נוסף - בו הוספנו את החישוב של המיקום הגיאוגרפי -City_mean_zacaut_geo_with_nafa.xlsx וחישבנו את המיקום הגיאוגרפי של כל יישוב/עיר על ידי חישוב ממוצע של המיקומים של בתי הספר הנתונים.
התמודדנו עם פערים בהשוואה (חוסר קונסיסטנטיות בשמות של כל המועצות אזוריות בארץ וערים עם שמות המורכבים ממספר מילים) באמצעות תיוג מקום ע"ב אתר https://geohack.toolforge.org/ ולפי הרישום בוויקיפדיה, לצד השלמת פערים ידנית נקודתית.
ליצירת הקובץ הסופי - city_mean_zacaut_geo_with_nafa_clean.xlsx

d. **בדיקת מהימנות:**

- בחינה למול נתונים כללים המתפרסמים בתקשורת – דוגמת העובדה כי ביהודה ושומרון (לפי אתרי חדשות) התקצוב לכל תלמיד הוא הגבוה ביותר בארץ, ואכן ניתן לראות כי באופן עקבי אחוז הזכאות בנפה זאת גבוהה מאוד ולרוב מקסימלית ברמה הארצית.
- מצדו השני של המתרס, באר שבע, צפת ואזור כנרת נפות עם אחוז זכאות נמוך יותר. לדוגמא נפת באר שבע מתפרשת על חצי משטחה של מדינת ישראל וכוללת אזורי פריפריה ואוכלוסיות מיעוט מוחלשות.
- בנפות המרכזיות, השינויים משמעותית קטנים יותר (אוכלוסייה יותר גדולה, קשה יותר 'להזיז' את הממוצע, כמצופה). לדוגמא - נפת תל אביב, ירושלים, באר שבע, לעומת צפת, רמת הגולן

e. **תוצרים ומסקנות** (המאגר המחושב מהווה תוצר בפני עצמו, לאחר שיפור ויזואליזציה ע"ב סקאלת צבעים כמפת חום):

i. ישנם הבדלים משמעותיים בין הנפות השונות שלרוב מצביעים על פערים הקשורים בפריפריה (חברתית או גיאוגרפית).

ii. ניתן לראות כי יש מגמה ארצית של שיפור בזכאות לבגרות לפי נפה כאשר באזורים מסויימים חל שיפור ניכר (צפת לדוגמא). לפי נתונים שראינו (ראו פר בית ספר) אכן יש מגמת שיפור, אם זאת ישנם בתי ספר שלא נכנסו לקריטריון זכאות לבגרות ולכן יש להסתכל על הנתונים בעין ביקורתית.

תצלום מסך מתוך zacaut_per_nafa_mean.xlsx, sheet: zacaut:

זכאות לבגרות ישראלית לפי נפות ושנים						
	תשעד	תשעה	תשעו	תשעז	תשעח	AVERAGE
חדרה	61.6273	60.3447	61.8702	68.342	69.8396	64.4048
יזרעאל	60.6717	63.6254	63.6161	67.1938	68.9303	64.8075
באר שבע	62.8011	64.5086	63.1223	66.2173	68.5072	65.0313
כנרת	63.4765	67.8438	62.2263	69.5278	66.0632	65.8275
צפת	57.8889	62.9556	66.3333	67.2111	74.75	65.8278
ירושלים	64.7065	65.65	66.2436	64.4211	71.7829	66.5608
השרון	65.2154	65.5684	67.6775	69.378	72.7488	68.1176
עכו	68.435	69.9675	68.8048	70.2286	69.0035	69.2879
אשקלון	64.4224	69.2018	70.7169	72.04	72.3185	69.7399
רמת הגולן	65.2625	64.1778	73.59	72.7182	78.975	70.9447
חיפה	71.0583	70.242	70.6176	73.1077	73.6833	71.7418
פתח תקווה	70.2136	71.1706	76.8721	74.9644	75.0026	73.6447
תל אביב - יפו	72.6317	73.3204	73.2374	74.5574	74.756	73.7006
רמלה	70.2514	70.4605	74.7525	76.48	79.4667	74.2822
רחובות	73.5229	76.3462	79.3075	80.2482	82.2158	78.3281
יהודה ושומרון	76.0314	76.5453	77.5786	80.815	83.0127	78.7966

iii. מבדיקת מגמת השיפור (בין תשע"ד לתשע"ח באחוזים) ניתן לראות מגמות שונות, לחיוב ולשלילה. בהתאם, ניתן להסיק מכך המלצות לקבלת החלטות ומדיניות – תוך העמקה איכותנית ברשויות שהשתפרו משמעותית ובסיבות לכך (דוגמת צפת) וגיבוש מדיניות לסיוע לרשויות בהן אחוז השיפור נמוך (דוגמת עכו).

iv. תצלום מסך מתוך zacaut_per_nafa_mean.xlsx , sheet: improvement:

שיפור בזכאות לבגרות באחוזים מתשע"ד לתשע"ח בחלוקה לנפות		
	שיפור בזכאות לבגרות באחוזים מתשעד לתשעח	אחוז זכאות תשעח
עכו	0.830698286	69.00348837
תל אביב - יפו	2.924729322	74.7559633
חיפה	3.694148	73.68333333
כנרת	4.075033288	66.06315789
פתח תקווה	6.820556927	75.0025974
באר שבע	9.085950449	68.50721649
יהודה ושומרון	9.182164717	83.01269841
ירושלים	10.93604665	71.78285714
השרון	11.55164941	72.74883721
רחובות	11.82335141	82.21578947
אשקלון	12.25667789	72.31846154
רמלה	13.1175099	79.46666667
חדרה	13.325837	69.83962264
יזרעאל	13.61201499	68.93030303
רמת הגולן	21.01130052	78.975
צפת	29.12667946	74.75

v. בבחינת הנתונים על ערים (מידע מלא ניתן למצוא בתוצר בקובץ city_mean_zacaut_geo_with_nafa_clean.xlsx בsheet = zacaut_by_year): נראתה מגמה דומה לנפות, של עליה באחוזי הזכאות לבגרות.

vi. בראש טבלת הערים מופיעה במפתיע בית ג'ן – השייכת לנפת עכו (אחת הנפות החלשות במדינה). לצדה ישנן כמובן גם ערים מחתך סוציו-אקונומי גבוה כצפוי (גני תקווה, גבעתים, השרון, גבעת שמואל, מיתר,שוהם).

- **מתוך טבלת זכאות לבגרות לפי ערים:**

idxCol	city	Nafa	geo_y	geo_x	תשעד	תשעה	תשעז	תשעו	תשעח	average
מגילות ים המלח	מגילות ים המלח	יהודה ושומרון	31.773735	35.503111	25	15.8	60	31.3	25	31.42
עין מאה'ל	עין מאה'ל	יזרעאל	32.722572	35.35293	31.05	32.7	27.15	34.85	34.9	32.13
אל קסום	אל קסום	באר שבע	31.246442	34.804604	47.86667	40.53333	29.975	37.5	35.775	38.33
נווה מדבר	נווה מדבר	באר שבע	31.245097	34.796603	37.3	38.5	39.225	33.05	46.65	38.945
טובא-זנגריה	טובא-זנגריה	צפת	32.96621	35.592018	28.7	30.5	41.7	37.9	58.2	39.4
מודיעין עילית	מודיעין עילית	יהודה ושומרון	31.934401	35.0418755		20	56.1	15.4	72.7	41.05
מג'ד אל-כרום	מג'ד אל-כרום	עכו	32.921203	35.254476	40.4	45.65	33.26667	43.95	42.53333	41.16
תל שבע	תל שבע	באר שבע	31.24597	34.85824133	48.26667	46.13333	39.1	35.33333	42.96667	42.36
ג'סר א-זרקא	ג'סר א-זרקא	חדרה	32.536655	34.912128	35.2	23.4	58.1	36.6	62.8	43.22
עילוט	עילוט	יזרעאל	32.717249	35.261482	29.8	30.3	58	28.6	73.7	44.08
בסמת טבעון	בסמת טבעון	יזרעאל	32.736287	35.1532	33.7	38.2	50	41	58.1	44.2
שפרעם	שפרעם	עכו	32.806063	35.17177	48.93333	42.87143	51.23333	40.67143	44.87143	45.71619
ביר אל-מכסור	ביר אל-מכסור	עכו	32.777291	35.220571	51.5	51.7	39.7	42	48.1	46.6
אופקים	אופקים	באר שבע	31.311436	34.6181034	43	47.14	44.6	45.96	52.6	46.66
שגב שלום	שגב שלום	באר שבע	31.198071	34.83931	46.05	37.65	57.7	45.7	50.76667	47.57333
מעלה עירון	מעלה עירון	חדרה	32.5487265	35.17799	34.3	43.7	57.25	47.6	58.05	48.18
אבו גוש	אבו גוש	ירושלים	31.804176	35.157816	47.5	40.7	43.6	50.9	59.8	48.5
אכסאל	אכסאל	יזרעאל	32.682085	35.320743	50.2	44.3	44.5	53.6	51.3	48.78
זרזיר	זרזיר	יזרעאל	32.690807	35.19609	31.8	41.7	56.1	56	61.1	49.34
זבולון	זבולון	חיפה	32.79323	35.117063	33.2	53.075	51.05	56.075	54.825	49.645
יואב	יואב	אשקלון	31.758651	34.786412	43	44.95	44.3	55.8	60.35	49.68
גן רוה	גן רוה	רחובות	31.917039	34.770069	38.3	44.7	67.2	39.2	59.7	49.82
מרום הגליל	מרום הגליל	צפת	32.970373	35.093739	33.9	56.9	55.5	38.7	65.75	50.15
רהט	רהט	באר שבע	31.3928965	34.7554	41.94	44.76	57.65	46.63333	61.25	50.44667
כסיפה	כסיפה	באר שבע	31.249246	35.083681	51.35	46.6	49.86667	54.4	54.96667	51.43667
רכסים	רכסים	חיפה	32.7505215	35.0988175	30	43.3	65.7	62.5	57.85	51.87
בני שמעון	בני שמעון	באר שבע	31.442291	34.761035	37.6	51.3	53.55	59.45	58.55	52.09
קלנסווה	קלנסווה	השרון	32.285791	34.981487	36	36.7	60.3	53.63333	74.35	52.19667
טירה	טירה	השרון	32.235466	34.950506	46.86667	61.15	52.66667	45.66667	54.86667	52.24333
בית אל	בית אל	יהודה ושומרון	31.9416	35.222734	86.05	87.76667	86.86667	86.8	87.76667	87.05
אשכול	אשכול	באר שבע	31.302878	34.43135	87.95	80.93333	86.775	87.5	92.9	87.21167
אפרת	אפרת	יהודה ושומרון	31.65836633	35.15619533	80.93333	86.23333	90.06667	88.36667	93.3	87.78
מנשה	מנשה	חדרה	32.467136	35.012976	82.725	87.025	90.3	88.475	94.45	88.595
רמת נגב	רמת נגב	באר שבע	31.00441	34.770368	84.9	91.4	79.7	98.9	89.5	88.88
בוקעאתה	בוקעאתה	רמת הגולן	33.201655	35.778757					88.9	88.9
רמת השרון	רמת השרון	תל אביב - יפו	32.13915775	34.83804875	84.6	89.725	88.875	88.9	92.525	88.925
קצרין	קצרין	רמת הגולן	32.993027	35.689823	89.6	88.3	91.25	91.45	85.75	89.27
פקיעין - בוקייעה	פקיעין - בוקייעה	עכו	32.97309	35.326711	92.7	89.9	87.1	87.5	90.4	89.52
מזכרת בתיה	מזכרת בתיה	רחובות	31.849076	34.839379	91.4	91.3	91.4	88.6	85.9	89.72
שפיר	שפיר	אשקלון	31.697167	34.728724	76.8	91.65	95.93333	90.2	95	89.91667
שוהם	שוהם	פתח תקווה	32.003884	34.947397	89.5	87.1	92.1	90.8	91.55	90.21
אלקנה	אלקנה	יהודה ושומרון	32.111538	35.0344335	84	85.35	96.6	89.8	97.85	90.72
מיתר	מיתר	באר שבע	31.327162	34.941613			91.1		91.9	91.5
גזר	גזר	רמלה	31.88859	34.917376	89.76667	90.4	92.63333	94.33333	90.6	91.54667
תל מונד	תל מונד	השרון	32.256138	34.921469	92.3	90.1	92.1	96.4	90.8	92.34
חורפיש	חורפיש	עכו	33.016688	35.347106	94.8	92	94.9	91	93.4	93.22
נס ציונה	נס ציונה	רחובות	31.93143633	34.79887533	91.3	91.9	92.3	95.53333	95.23333	93.25333
לב השרון	לב השרון	השרון	32.260652	34.894631	90.9	92	94	94.6	95.5	93.4
יבנה	יבנה	רחובות	31.87448625	34.746518	93.05	97.78	89.92	94.6	92.8	93.63
דרום השרון	דרום השרון	פתח תקווה	32.133649	34.911086	91.95	92.5	96.35	96.45	91.46667	93.74333
כאוכב אבו אל-היג'א	כאוכב אבו אל-היג'א	עכו	32.830676	35.248658		98.3	85.3	96.7	96.3	94.15
גדרה	גדרה	רחובות	31.811044	34.78039175	90.125	92.475	97.05	95.075	96.4	94.225
גבעת שמואל	גבעת שמואל	פתח תקווה	32.076281	34.8502555	95.5	92.7	93.5	93.6	97.6	94.58
גבעתיים	גבעתיים	תל אביב - יפו	32.0726485	34.80892975	93.8	93.225	96.775	93.275	97.025	94.82
גני תקוה	גני תקוה	פתח תקווה	32.059893	34.87532		100	90	100	90	95
גוש עציון	גוש עציון	יהודה ושומרון	31.658085	35.118082	92.725	94.7	96.55	94.9	98.375	95.45
קרית עקרון	קרית עקרון	רחובות	31.86169	34.822476	94.5	98.2	98.2	93.2	96.1	96.04
בית ג'ן	בית ג'ן	עכו	32.964282	35.378468	100	98.5	99	99.5	92.6	97.92

4. **פערים בזכאות לבגרות בתוך ערים:**
   a. **קובץ קוד:** - GapPerCityNafa.py
   b. **מטרה:** בחינת פערים בתוך עיר בין ביה"ס בעלי זכאות מקסימאלית ומינימאלית, כמדד לאי-שיוויון בתוך רשות מוניציפאלית המחייבת בחינה והעמקה.
   c. **הסבר התהליך:**
      i. עבור כל אחת מהשנים המדוברות חישבנו לכל עיר את ההפרש בין בית הספר עם אחוז הזכאות המקסימלי לבין בתי הספר עם אחוז הזכאות המינימלי ושמרנו בקובץ עזר city_gap_zacaut_with_nafa.xlsx. לאחר מכן היה צורך בתיקונים מינורים למול בחינה של האקסל, הוספנו עמודת ממוצע ולשם הנוחות הויזואלית צבענו את הערכים בסקלאת צבעים ומיינו לפי עמודת הממוצע לבחינת דפוסים.

d. **תוצר**:

i. לאחר מכן איחדנו את הנתונים לקובץ סופי ב city_mean_zacaut_geo_with_nafa_clean.xlsx בsheet =zacaut_gap שם הוספת עמודת ממוצע וצבענו לפי סקאלת צבעים של מפת חום (מהפער הקטן והטוב לפערים הגדולים). לאחר אפיון התוצאות החלטנו להשמיט ישובים בעלי פער בגודל אפס (לאחר בדיקה מדגמית כי מדובר בישובים בהם יש בי"ס בודד).

ii. התוצר מונגש גם כשכבה על מפה (שכבה מס' 10).

5. בחינת כמות השקעה בתלמיד בכל בית ספר ובכל עיר:

a. **קובץ קוד:** Equality.py

b. **מטרה:** אפיון הפערים בתקציב לכל תלמיד בבתי הספר תחת ההנחה כי הדבר מהווה פוטנציאל השפעה מרכזי גם על הישגי התלמידים.

c. **הסבר התהליך:**

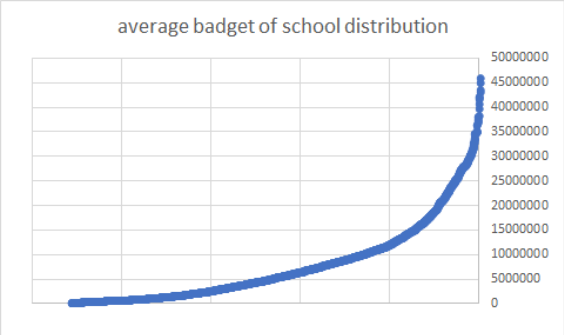
i. יצרנו בחינה השוואתית של ההשקעה בכל תלמיד מתוך התקציב הכללי, ולאחר מכן בחנו את הנתונים בבתי ספר באותה העיר. כך ניתן לבחון הן את ההבדלים בין בתי הספר ובין הערים והן לנסות להצביע על מרכיבים המשפיע על פערים אלו.

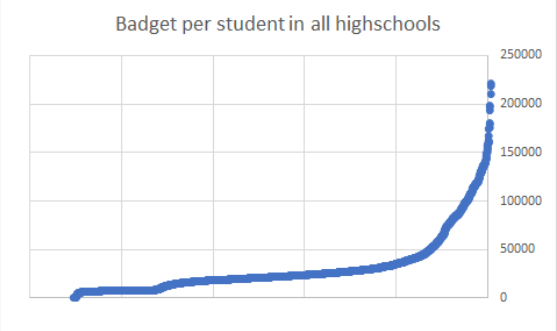
d. **תוצר ומסקנות:**

i. ניתן לראות שביה"ס בהבדלים משמעותיים, שנובעים בחלקם מגודל בתי הספר ומאפייניהם (היקף החינוך המיוחד בביה"ס לדוגמה), שמשפיע למעשה בסופו של דבר על כמות ההשקעה התקציבית עבור כל תלמיד.

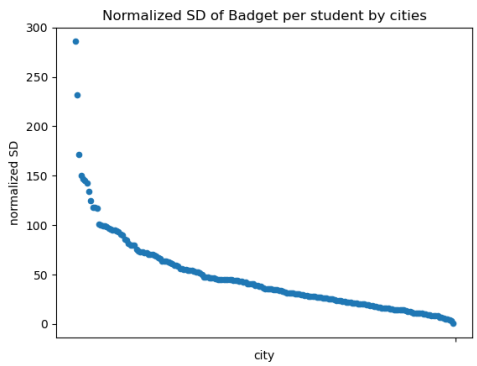
ii. בבחינת בתי הספר, ניתן לראות שההבדלים בין התקציב לתלמיד בכל בית ספר הם במגמה מעריכית. מבדיקה מדגמית של קצוות ההתפלגות, ניתן לראות את ההשפעה של גודל בית הספר על מאפייני התקציב לכל תלמיד ובכך, שחלק מהתקציב של בתי הספר מושקע בעלויות קבועות.

iii. כמו כן, מבדיקת התפלגות ממוצע התקציב (ללא התחשבות בכמות התלמידים) ניתן לראות מגמה דומה שבה הנתונים מפולגים כמובן באופן שונה.





iv. כמו כן, בבדיקה משוקללת על כל עיר (ע"ב סטיית תקן מנורמלת לסכום התקציב) ניתן לראות כי ישנן ערים עם הבדלים משמעותיים בין בתי הספר שלהן בתקציב לכל תלמיד, דבר שיכול לשמש נתון חשוב עבור הפרט בקבלת החלטות להצטרפות לבי"ס, וכן כסוגייה לבחינה עבור מקבלי ההחלטות.



Normalized SD	City
285.848810	בנימינה גבעת עדה
231.896647	לב השרון
171.361717	כסרא סמיע
150.353198	גבעת זאב
146.979120	הגליל התחתון
....	
5.218368	כפר יונה
4.903493	מזכרת בתיה
4.364335	אשכול
3.116830	ברנר
0.475451	בענה

בתי ספר בעלי תקציבי קיצון (גבוה ביותר ונמוך ביותר לכל תלמיד):

per student	City	School name	School
220848.7	ירושלים	המגשימים	358127
218266.5	בנימינה גבעת עדה	הרים	353706
209832.8	הגליל התחתון	בנימין רוטמן כדורי	512491
198231.8	קריית שמונה	גוונים	620385
197215.4	מודיעין-מכבים-רעות	פסיפס	731182
...			
149.2963448	ראשון לציון	אורט למינהל טכני	900126
49.51419686	ירושלים	אליתים אלערבי	378125
22.45297428	ירושלים	ישיבת פאר יעקב	363150
22.33630769	מטה יהודה (גבעת שמש)	הילה רטורנו-חסות	961011
16.84117647	ירושלים	אורה ושמחה	166538

6. בחינת מאפייני חלוקת בתי הספר

a. **קובץ קוד:** zacaut.py

b. **תצורת עבודה מרכזית:** ניתוח השוואתי ותיאורי של התפלגות הנתונים.

c. **מטרה:** השוואה בין פילוחים שונים של בתי הספר לתתי קבוצות, למציאת מודלי חלוקה המגדירים בצורה טובה תתי קבוצות, מתוך הבנה כי אופן החלוקה משפיע על תצורת העבודה הרוחבית של מערכת החינוך ודרך כך גם על תקציבים ומשאבים.

d. **הסבר התהליך:**

i. ע"ב טבלת "school – bagrut.xls" בחנו את ההתפלגות של הזכאויות השונות לבגרות לפי שני מודלים של קבוצת דומים מאותם סדרי גודל (20 במודל שלנו ו27 במודל הקיים), החופפים באופן חלקי:

1. במודל הקיים – כלומר, חילצנו קבוצת דומים מנתוני PROXY שלהם (ממוצעי זכאויות לפי קבוצת דומים).
2. בחלוקה לתתי קבוצות של כל הצטלבות מגזר ואחוזון טיפוח סוציואקונומי.

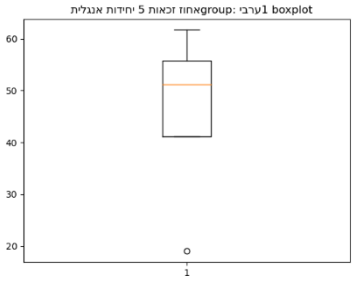
ii. ניתחנו מדדי שונות, ונעזרנו בתוצרים גרפיים של Box Plots, במטרה להציגם ככלי ניהולי לאפיון תתי קבוצות (ממוצעים, ערכי קיצון וכו') ולגיבוש מדיניות תואמת.

iii. במקביל, בחנו את אופן התפלגות הנתונים לפי החלוקה שהצענו ביחס לנתונים הקיימים – כהוכחת ערכיות לחלוקה זו עבור חלק מתתי הקבוצות, בניסיון להקטין שונות – כמדד שמעיד על מדיניות אחודה שאינה תואמת את כלל בתי הספר בקבוצה.

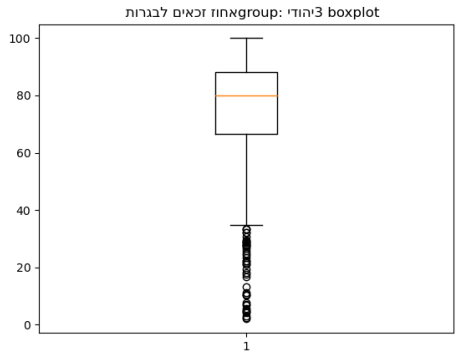
e. **תוצר ומסקנות:**

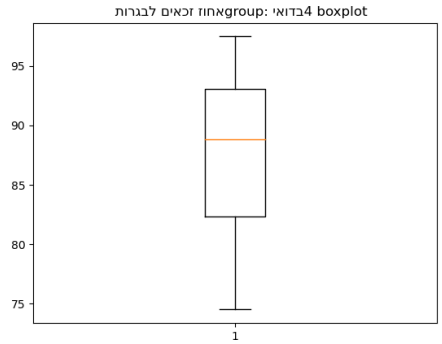
i. מצאנו כי בבחינת הזכאות ל5 יחידות מתמטיקה, 5 יחידות אנגלית ובגרות בהצטיינות, קיימות תתי קבוצות בעלות שונות נמוכה יותר (לדוגמה בקבוצת אחוזון 1 במגזר הערבי החופפת ברובה לתת קבוצה קיימת):

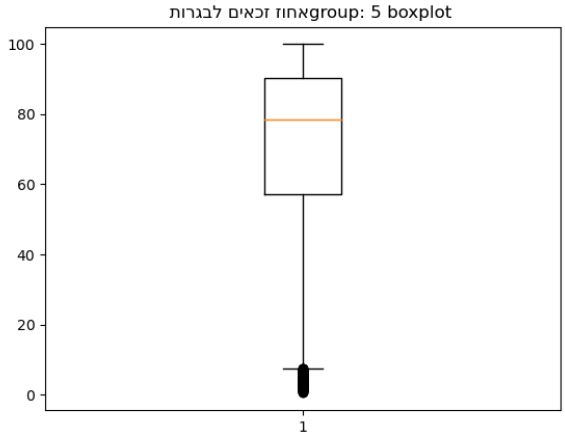
Var	mean	zacaut kind	חלוקה
410.9903	45.94624	אחוז זכאות 5 יחידות אנגלית	קיימת
347.6292	45.775	אחוז זכאות 5 יחידות אנגלית	1ערבי



ii. בקבוצה קיימת המכילה ערבוב בין אחוזון 3 היהודי לבין אחוזון 4 הבדואי, יש פער משמעותי בין הישגי בתי הספר היהודים לאלו הבדואיים – כלומר ניתן לראות את "המרוויח והמפסיד" מאותה תת חלוקה, ככלי לקבלת החלטות ניהוליות. (להלן ה-BoxPlots של תתי הקבוצות שלנו, השייכות לאותה "קבוצת דומים" קיימת (מס' 5) המופיעה בגרף מתחת.







iii. מבדיקת מקרה בוחן, להשוואה נקודתית בין קבוצות דומים קיימות לבין הנתונים המקבילים להן בקבוצות החלוקה שלנו, כך לדוגמה: עבור קבוצת הדומים בה ממוצע הזכאות לבגרות הינו 56.9 עולה כי היא מאופיינת כולה באחוזון ['5'] ובמגזרים ['יהודי', 'בדואי', 'ערבי', 'דרוזי'], וכך-כזו השונות בה הינה: 356.192330881591.
בהשוואה לחלוקה שלנו, ניתן לראות בפירוש שתת קבוצה זו מתפלגת בצורה לא אחידה הן בשונות תתי הקבוצות והן בממוצעים, כך שהיא עשויה לגרום לפגיעה בבתי ספר בשולי החלוקה.

quintile	sector	zacaut kind	Mean	Var
5	יהודי	אחוז זכאים לבגרות	60.00433	569.4053
5	בדואי	אחוז זכאים לבגרות	50.46413	220.1019
5	ערבי	אחוז זכאים לבגרות	60.66691	420.9802
5	דרוזי	אחוז זכאים לבגרות	77.58864	108.9801

תוצר נתוני הזכאות לבגרות בהשוואה בין החלוקה לפי תתי קבוצות מגזר ואחוזון טיפוח לעומת החלוקה הקיימת :

אחוזון טיפוח ומגזר				חלוקה קיימת ל"קבוצות דומים"		
quintile	sector	Mean	Var	#	mean	var
5	יהודי	60.00433	569.4053	1	22.40633	378.0855
5	בדואי	50.46413	220.1019	2	25.38506	513.5985
5	ערבי	60.66691	420.9802	3	26.744	470.2158
5	דרוזי	77.58864	108.9801	4	31.92857	526.9009
נוער בסיכון	יהודי	30.6102	524.8423	5	32.42593	482.3055
נוער בסיכון	בדואי	23.6	246.42	6	56.02778	447.7768
נוער בסיכון	ערבי	17.34405	184.3697	7	56.355	486.1028
נוער בסיכון	דרוזי	4.2	-	8	58.25743	405.4378
4	יהודי	68.58438	454.6857	9	60.65068	401.991
4	בדואי	87.4	73.376	10	62.9973	428.215
4	ערבי	68.87017	337.9727	11	64.37871	446.2432
4	דרוזי	79.33243	196.1278	12	67.49049	459.341
3	יהודי	74.51467	423.9713	13	71.40598	411.1913
3	ערבי	80.07222	176.2229	14	71.67958	464.4517
3	דרוזי	69.2	-	15	74.42667	293.772
2	יהודי	80.01541	343.9218	16	74.61277	304.5173
2	ערבי	91.43714	78.09512	17	74.85419	371.2916
2	דרוזי	99.1	1.73	18	76.56447	466.6575
1	יהודי	86.73019	239.8388	19	77.37645	382.588
1	ערבי	95.375	52.61583	20	80.44337	381.1697
				21	82.75329	279.5271
				22	84.11098	230.0887
				23	85.61726	253.4089
				24	86.12049	237.6285
				25	86.24455	284.6688
				26	86.79561	236.8335
				27	88.8963	184.3151

7. קלסטרינג לפי אחוזי בגרויות

a. **קובץ קוד:** cluster for schools.py

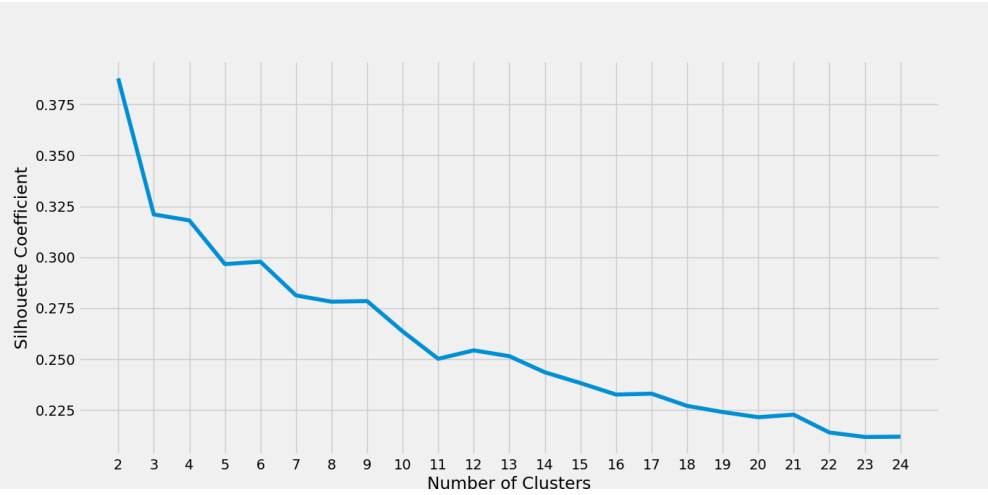
b. **תצורת עבודה מרכזית:** שימוש בקלאסטרים במודל K-Means.

c. **מטרה:** השוואה בראי התוצאה – כלומר, השוואה בין בתי ספר מרקעים שונים החולקים את אותו מרחב הישגים, וכן לבחון למול חלוקות למגזרים ואחוזוני טיפוח.

d. **הסבר התהליך:**

i. התבססנו על k-means מתוך ניסיון לבחון את פילוח הקבוצות בסדרי גודל ידועים, מתוך הנחה שיישום של השיטה בפועל ידרוש תפעולית עבודה במספר קבוצות בסדר גודל מסוים. את הקלאסטרים יצרנו על בסיס כל נתוני הזכאות לבגרות השונים (כללי, יח"ל במתמטיקה ואנגלית, הצטיינות וכו') ובכך ייצרנו תתי-קבוצות בעלי מאפיינים יחודיים שלא ניתן לחלק באמצעות פילוח קטגוריאלי פשוט.

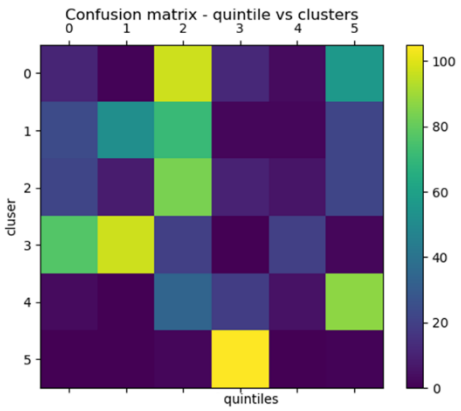
ii. בהתאם, לאחר בדיקה של הSSE וsihloutte התמקדנו ב3 חלוקות: {6,17,21} ובחינת התפלגותם למאפייני בתי הספר (מגזר, אחוזוני טיפוח, סוג פיקוח וכו') ע"ב טבלת "school – bagrut.xls".

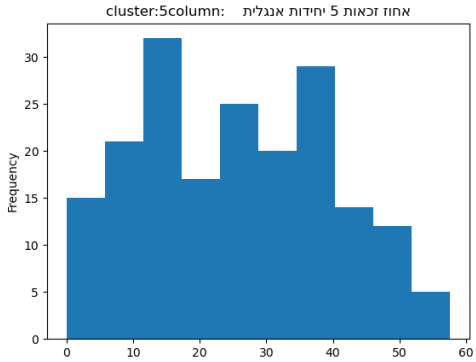


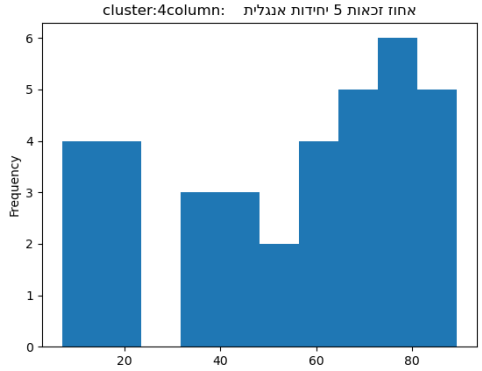
iii. נעזרנו בכלי Confusion Matrix להשוואה בין הקלאסטרים לבין חלוקות מאותם סדרי גודל, לצד פלט סטטיסטיקה תיאורית על כל קלאסטר וכל תצוגה בהיסטוגרמות של מאפייני בתי הספר בכל קלאסטר.

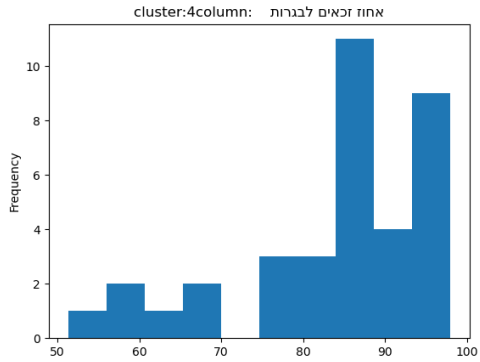
e. **תוצר ומסקנות:**

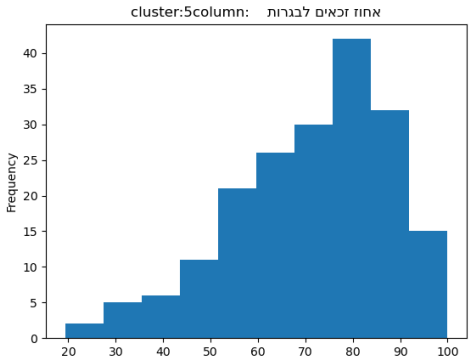
i. כך לדוגמה, בבחינת החלוקה ל6 קלאסטרים, השוואנו את הקלאסטרים לאחוזוני הטיפוח (הממוספרים 1-5, וכן אחוזון נוער בסיכון סומן כ– 0), וניתן לראות כי אחוזון טיפוח 3 מחולק בין רוב המצוי בקלאסטר 5 לבין מיעוט הנמצא בקלאסטר 4 (ומיעוט בנוספים). כך, בהסתכלות על הישגיהם השונים – ניתן להבנות תוכניות יחודיות לבתי ספר מאותו אחוזון טיפוח בהתאם לקבוצת השווים להם בהישגים – כך לדוגמה ניתן לראות את ההתפלגות השונה בזכאות ל5 יחידות אנגלית, וכן בהתפלגות הזכאות לבגרות באופן כללי.



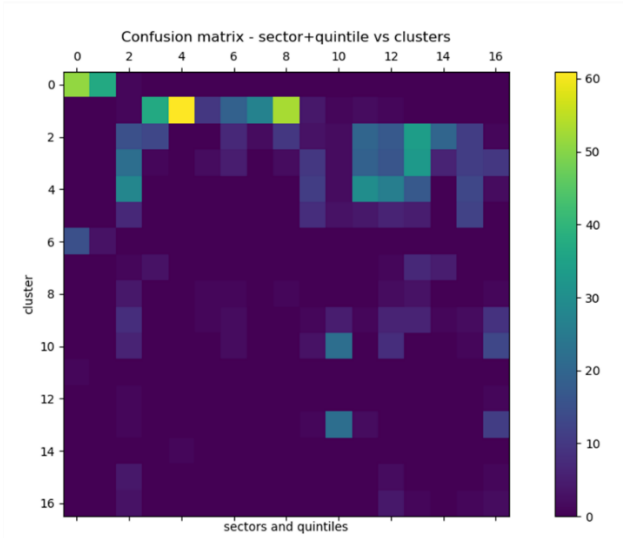






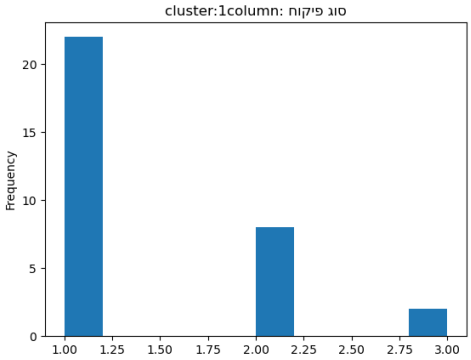


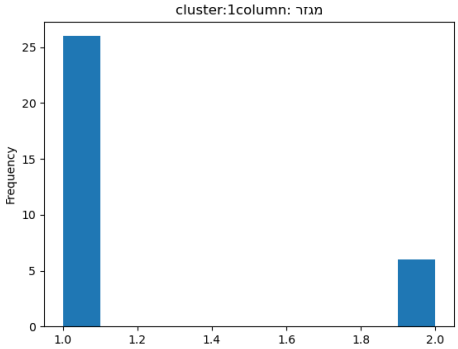
ii. דוגמה נוספת לשימוש זה ניתן לראות גם בהשוואה בין החלוקה שהגדרנו מעלה המשלבת מגזר ואחוזון טיפוח (לדוגמה: "אחוזון טיפוח 3 מגזר יהודי"), לקלסטרים של הזכאות לבגרות (מאותו סדר גודל). כך לדוגמה, בבחינת החלוקה ל6 קלאסטרים, השוואנו את הקלאסטרים לאחוזוני הטיפוח (הממוספרים 1-5, וכן אחוזון נוער בסיכון סומן כ– 0), וניתן לראות כי אחוזון טיפוח 3 מחולק בין רוב המצוי בקלאסטר 5 לבין מיעוט הנמצא בקלאסטר 4 (ומיעוט בנוספים). כך, בהסתכלות על הישגיהם השונים – ניתן להבנות תוכניות יחודיות לבתי ספר מאותו אחוזון טיפוח בהתאם לקבוצת השווים להם בהישגים – כך לדוגמה ניתן לראות את ההתפלגות השונה בזכאות ל5 יחידות אנגלית, וכן בהתפלגות הזכאות לבגרות באופן כללי.



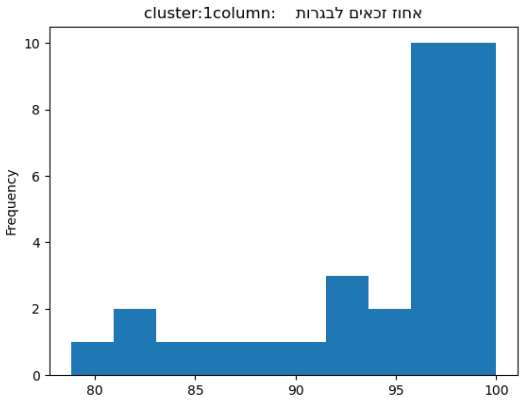
מקרא	
קבוצה	סימון
יהודי נוער בסיכון	0
יהודי אחוזון 1	1
יהודי אחוזון 2	2
יהודי אחוזון 3	3
יהודי אחוזון 4	4
יהודי אחוזון 5	5
ערבי נוער בסיכון	6
ערבי אחוזון 2	7
ערבי אחוזון 3	8
ערבי אחוזון 4	9
ערבי אחוזון 5	10
בדואי נוער בסיכון	11
בדואי אחוזון 4	12
בדואי אחוזון 5	13
דרוזי אחוזון 2	14
דרוזי אחוזון 4	15
דרוזי אחוזון 5	16

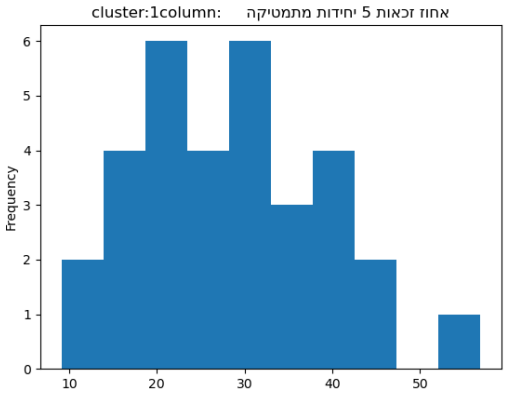
iii. ניתן לאפיין זיקה ברורה בחלק מהקבוצות למאפייני הישגים מסוימים, לדוגמה ניתן לראות כי בתי ספר יהודים מאחוזונים 3 ומטה חולקים את אותו קלאסטר הישגים עם בתי ספר ערבים באחוזונים 2-3, דבר שיאפשר שיתופי פעולה או פעילות של מערכת החינוך שלא היתה מתגבשת אחרת.
מהעמקה במקרה מבחן זה ניתן לראות כי מרבית בתי הספר המשויכים לקאסטר זה הם מהמגזר היהודי (כ-83%) וברובם הם בפיקוח ממלכתי דתי (סימול 1) המהווים כ-80% מהקלאסטר.

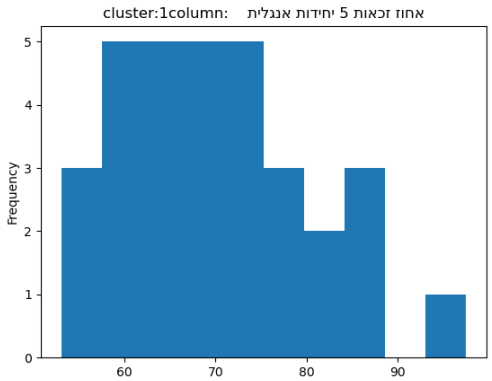


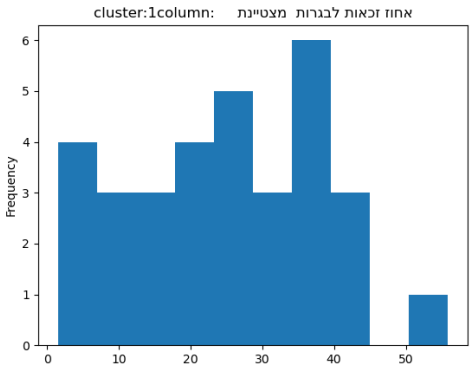


iv. בבחינת ההישגים המאפיינים את הקלאסטר, עולים אחוזי זכאות גבוהים לבגרות בכלל, ואחוזים גבוהים מהממוצע בפרט בבגרות בהצטיינות וב-5 יח"ל מתמטיקה ואנגלית. כלומר, מדובר בקבוצת בתי ספר חוצת מגזרים וסוג פיקוח (ערבי-ממלכתי לעומת ממלכתי-דתי יהודי).



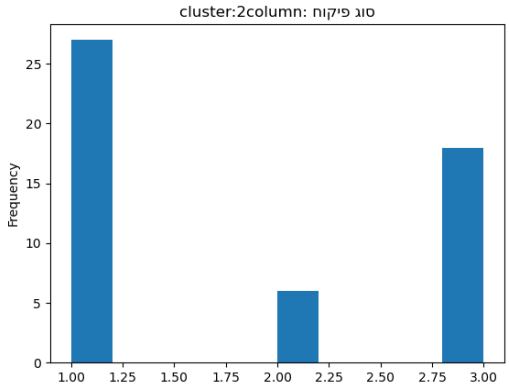


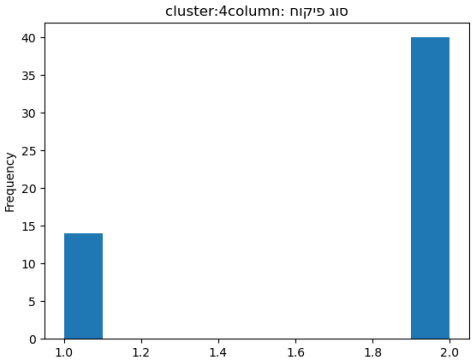




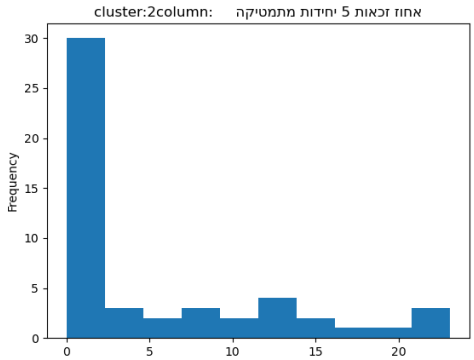
v. כמו כן, ניתן לראות כי מרבית בתי הספר הבדואים (קבוצות 11-13) מאופיינים באותם קלאסטרי הישגים בסיסיים עם בתי ספר יהודים מאחוזון טיפוח 2 אשר מבחינת המקרה עולה כי מרביתם ממלכתי דתי וחרדי. זאת כל הנראה בין היתר לאור פערים דומים בזכאות ל-5 יחידות במתמטיקה. כלומר ניתן לגבש מדיניות לקבוצה משותפת שלא חולקת כלל נתונים סוציואקונומיים ומגזריים.

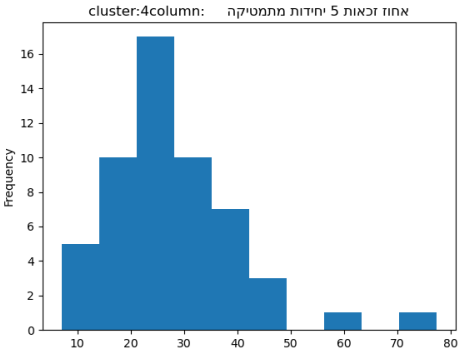
מאפייני בתי הספר בקלאסטרים 2 ו-4:





זכאות לבגרות ב-5 יח"ל במתמטיקה:





תוצרים אינטראקטיביים:

כפי שציינו מעלה, המרכיב השני בפרויקט הוא יצירת ממשקים אינטראקטיביים לניתוח והעמקה בתוצרים ובנתונים המעובדים.

1. **מפות - גוגל מפס**
   a. **תצורת עבודה מרכזית:** יצירת שכבות מידע ע"ב מפה.
   b. **קישור לתוצר:**
      i. **שכבות המידע הגיאוגרפיות בממשק Google maps-** https://www.google.com/maps/d/edit?mid=1OWSj1btTp1KNaz1a5FFvL_bEGCyyHv0r&usp=sharing
      ii. **סרטון הדגמה-** https://www.youtube.com/watch?v=Ro5ZRn4kIyk
   c. **מטרה:** שימוש בניתוחי הנתונים שביצענו ובמידע מנוקה, והנגשתם כשכבות לניתוח אקספלורטיבי ע"ב מפה.
   d. **הסבר על התהליך והתוצר:**
      i. יצרנו שני סוגים של שכבות מידע: שכבות הנגשת מידע רקע כללי – לצרכי סקירה והשוואה שכבות המידע המסוכם (שכבות 1-7); שכבות מבוססות ניתוח ונתונים אגרגטיביים בבחינת בתי ספר ועירים (שכבות 8-10). בכך:
         א. שכבה 1: תיכונים לפי נפה – ניתוח שייכות ביי"ס לנפה אזורית והצגה בחלוקה לצבעים (ע"ב קובץ מעובד -schools_data_only_hs.xlsx).

ב. שכבות 2-3 : תיכונים לפי מגזר (יהודי, ערבי, דרוזי ובדואי) וסוג פיקוח (ממלכתי, ממל״ד, חרדי), בהתאמה. (ע״ב קובץ מעובד - school_zacaut_nafa_jewish.xlsx).

ג. שכבות 4-5 : זכאות ל-5 יח״ל במתמטיקה ובאנגלית, בהתאמה, והצגת בתי הספר לפי סקאלת חום (גווני כחול מבהיר לכהה) (ע״ב קובץ - zacaut_tsah_full.xlsx).

ד. שכבות 6-7 : בתי ספר לפי היקף תקציב לכל תלמיד בתשע״ח, והיקף שעות לימוד פרטניות בתשע״ח, בהתאמה. בתי הספר צבועים לפי מפת חום (מבהיר לכהה). (ע״ב ניתוח מעובד של נתוני התקציב המובע בקובץ - per_stud_data_for_map.xls).

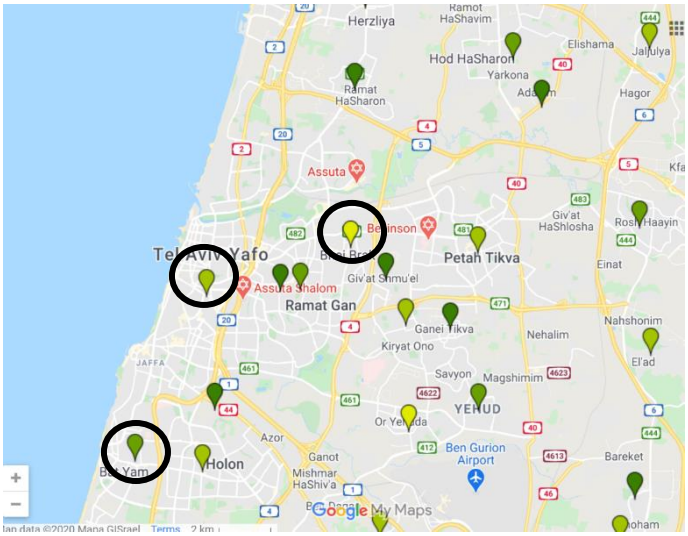
ה. שכבות 8-9 : ממוצעי זכאות כללית בכל עיר לשנת תשע״ח, וממוצע זכאות בערים בין השנים תשע״ד לתשע״ח, בהתאמה. צבועים כמפת חום (ירוק-אדום). (ע״ב קובץ מעובד - city_mean_zacaut_geo_with_nafa_clean.xlsx).

ו. שכבה 10 : מיפוי פערי זכאות לבגרות בין בתי ספר קיצוניים (מינ׳ מול מקס׳) באותה העיר בשנת תשע״ח. הערים צבועות ע״ב סקאלת מפת חום הפוכה (אדום – פער גדול, עד ירוק – פער נמוך). (ע״ב הקובץ בסעיף ה׳).

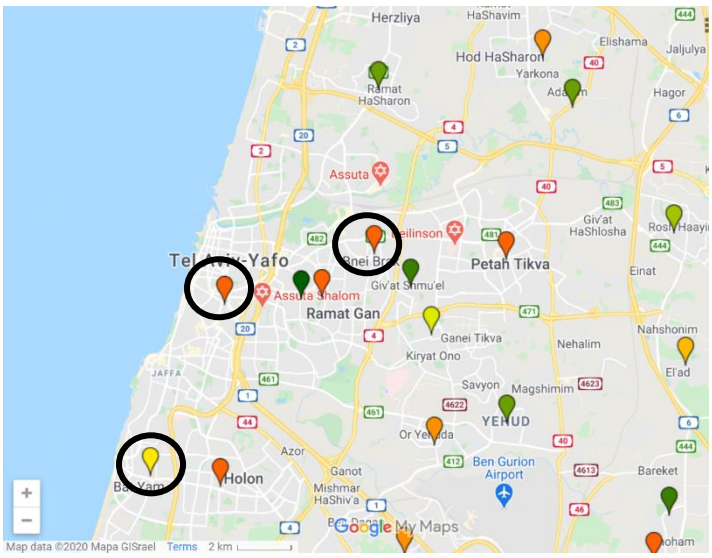
e. **מסקנות מהמפות:**

i. הבדלים בין ערים סמוכות – בבחינת ערים סמוכות בעלות פערי זכאות, ניתן לראות הבדלים גם בהיקף שעות הלימוד הפרטניות. כך לדוגמה : בהסתכלות על בת ים, תל אביב ובני ברק, ניתן לראות כי הממוצע הגבוה ביותר הינו דווקא בבת ים, לאחריו תל אביב ולבסוף בני ברק, והפערים בין בתי הספר בעיר גדלים באותו סדר בהתאמה והיקף שעות הלימוד הפרטניות גדל בהתאמה.

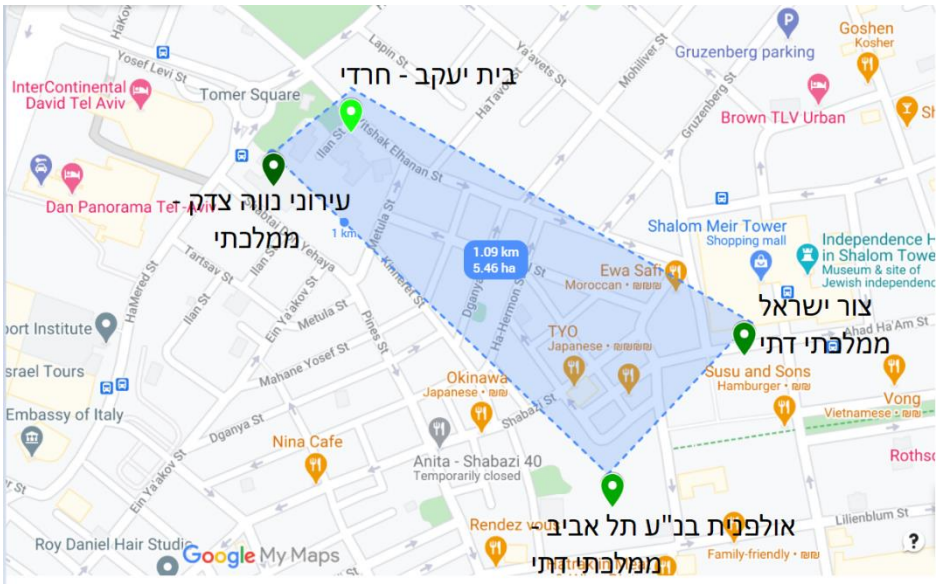
מפת ממוצעי הבגרות:



מפת הפערים בזכאות בין בתי ספר בעיר:

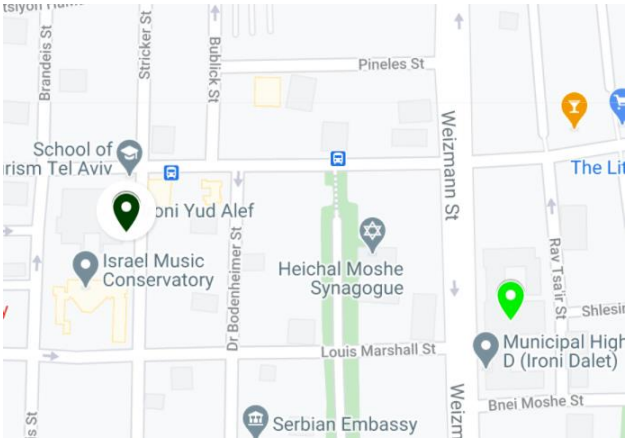


ii. הבדלים בתקצוב באותה עיר בין בתי ספר סמוכים– ע״ב המפה, בולטים הבדלים משמעותיים בתקציב לכל תלמיד בין בתי ספר סמוכים. כך לדוגמה, בתל אביב – מבדיקה על השכבות השונות, בולט כי בתי הספר ממערכת פיקוח שונה, ההבדלים בזכאות לבגרות מקבילים להבדלים בתקציב וכן להבדלים בהיקף שעות לימוד פרטניות (המושפע גם מהיות אחד בתי הספר מותאם לחינוך מיוחד). מקרה מנוגד, בולט בין תיכון י״א ועירוני ד׳ הסמוכים, בהם בולטים פערים משמעותיים בתקציב ובשעות הפרטניות במגמות הפוכות (מהעמקה- תיכון י״א מיועד לנוער מרקע קשה).

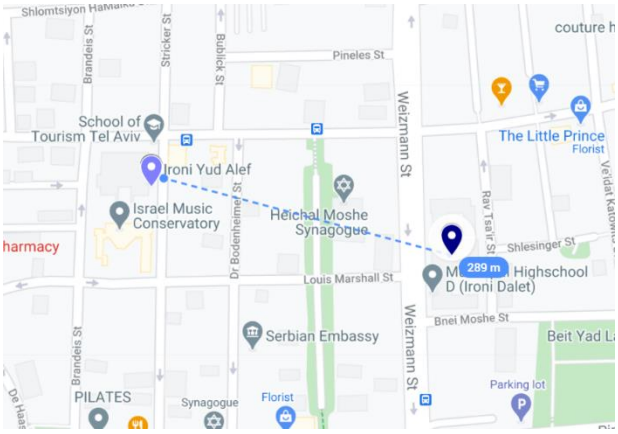


השוואה בין תיכון י״א לתיכון עירוני ד׳ (מרחק של 289 מ׳):

מפת תקצוב



מפת שעות פרטניות



2. **מודל לומד אקספלינאבילי כעזר למדיניות**

a. **תצורת עבודה מרכזית:** יצירת ממשק להפקת עץ החלטות.
b. **קישור לתוצר:**
   i. **קובץ פיית׳ון: DesTree (ע״ב קובץ גולמי school – bagrut.xlsx).**
   ii. **סרטון הדגמה-** https://youtu.be/dghC7nIH-ts
c. **מטרה:** ניתוח מאפייני התפלגות הזכאות לבגרות של בתי הספר, ממוקד ביעד זכאות אחד, ונועד לייצר ״מפת דרכים״ להבנת הגורמים הקורלטיביים להישגים שונים.
d. **הסבר על התהליך והתוצר:**

a. בחרנו להשתמש בעץ החלטות, כמודל הניתן להבנה (אקספלינאבילי) הלומד על המידע בשנה בודדת ומפלג את בתי הספר לפי נתוני השייכות שלהם לרשויות, מגזרים וצורות פיקוח שונות.

b. המודל מונגש באמצעות ממשק משתמש המקבל את הנתונים הרצויים – סוג זכאות רלוונטי, שנה לבחינת הנתונים, ומאפייני המדידה- תוך מתן אפשרות לשאול שאלה על ציון סף (לדוגמה – 60% ומעלה) או לפי חלוקה לתתי-טווחים קטגוריאליים (לדוגמה – 0-20,21-40 וכו'). המודל לומד על סט לאימון ומנגיש את רמת האמינות שלו (בבדיקה על סט נתונים נפרד לבחינה) ומציג לבסוף תוצר ויזואלי של עץ החלטות עם פרטי התנאים המבדלים. כך, מקבל החלטות יכול להשתמש בפרמטרים האלו כדי למצוא מוקדים מרכזיים של פערים (פיצ'רים המבדלים באופן מרכזי בין הלייבלים השונים של העץ) ולדייק את מדיניותו למול פערים אלו.

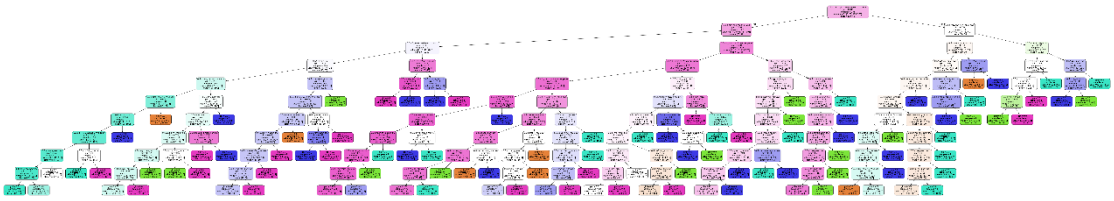
c. בחרנו לחלק את הדאטה בחלוקה של 75% לאימון, מתוך הנחה כי המודל משמש בעיקר לניתוח המציאות ולא לניבוי בפועל, והדאטה לבחינת התוצאות נועד לשמש כבדיקת מציאות של המודל שהמשתמש יוצר ולא מעבר לכך.

e. מסקנה לדוגמה:

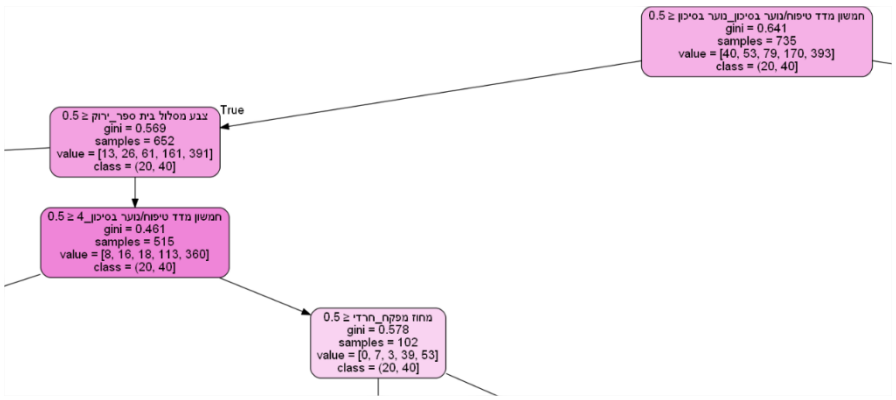
a. לאחר הפקת עץ זכאות כללית לשנת תשע"ח, בחלוקה לקבוצות טווח בקבוצות של 20%, התעמקנו בראש העץ בענפו השמאלי.

b. בולט כי שייכות לקבוצת נוער בסיכון היא המדד המשמעותי ביותר לפילוח, ומבין בתי הספר העונים לקריטריון זה, שייכות בי"ס ל"מסלול ירוק" (מדד של מערכת החינוך) היא המבחן המשמעותי אחריו.

c. כך יוכל האחראי במערכת החינוך להתמקד בבעיות המרכזיות בתחומו בהתאם למאפיינים השונים של בתי ספר – כלומר להתמקד בתת חלוקה המגדירה באופן משמעותי ביותר את האימפקט על שייכות לטווח מסוים של זכאות לבגרות אותו רוצים לתקן או שאליו רוצים לשאוף.



העמקה בראש השמאלי של העץ:



מכשולים -

במהלך העבודה נתקלנו בכמה מכשולים מרכזיים.

1. <u>ריבוי ופיזור המידע.</u> - המידע שעמד לרשותנו היה נתון בקבצים רבים, כל קובץ עם מדדים שונים, חלקים ברמה הבית ספרית וחלקים ברמה העירונית. בשלב הראשון, לפני שהתחלנו לנתח את המידע היה עלינו לסדר את אותו. יצרנו כמה קבצי בסיס המכילים את רוב הנתונים הרלוונטים (כל קובץ מורכב ממספר קבצים).
2. <u>התמודדות עם פערים במידע</u> - למרות שכל המקורות שהיו בידינו נלקחו ממקורות רשמיים, להפתעתנו הייתה חפיפה חלקית בלבד בין הקבצים השונים. התחלנו את הפרוייקט כשאנו בוחנים קרוב לעשרת אלפים שורות (כל שורה מייצגת בית ספר בשנה מסויימת). התצמצמנו לכמעט אלף בתי ספר רלוונטים (תיכונים) שעליהם היה לנו מידע בקבצים השונים. היה צורך בתיקונים ובדיקה ידנית, שכן הבדלים ברישום (שגיאות כתיב, שמות מתורגמים מערבית, כתיב מלא וחסר, התמודדות עם רווחים) גרמו לפערים בטבלאות המאוחדות.
3. <u>מציאת הכלים המתאימים לניתוח המידע והצגתו -</u> אחד הקשיים המרכזיים שלנו היה להבין מה הכלים המתאימים לניתוח המידע. היה עלינו ללמוד אלגוריתמים לומדים, אלגוריתמים של קלסטור, שימוש בAPI של גוגל מפות וספריות שונות לשליפת מיקומים (להשלמת מידע חסר).
4. <u>התאמה של מידע קטגורי לניתוחים סטטיסטים כמותיים/ למודלים של MECHINE LEARNING -</u>

- בקלאסטרינג היה צריך לסדר את המידע על מנת שהאלגוריתם יוכל לקרוא את הנתונים ולהשוות בין פילוח קטגורי (לדוגמא לפי מגזרים) לעומת פילוח לפי נתונים רציפים (לדוגמא אחוז זכאות לבגרות).

- בעץ החלטות - הפיכת נתון קטגורי לבינארי על ידי ייצור מאפייני Dummie והפיכת נתונים קטגורים לרציפים בדומה לקלסטרינג (לדוגמא עץ ההחלטות שלנו מתבסס אומנם על כ 8 עמודות אך על יותר מ200 פיצ'רים, היות וכמעט על לכל עמודה ייצרנו מאפיין Dummie).

עבודה אפשרית עתידית -

הפרויקט מהווה הוכחת היתכנות לשימושים קבועים במידע הזה בצורה נוחה וויזואלית, בהתאם - עבודה עתידית מרכזית יכולה להיות על נתונים עדכניים מהשנים האחרונות (2018-2020) ובחינת המודלים גם עליהם.

כמו כן, ניתן לבצע העמקות נוספות הממוקדות במאפייני תקציב כרכיב הנמדד - לבחון קלאסטרים ועץ החלטות על בסיסים (כלומר להתמקד בנתוני המוצא של בתי הספר ולא בהישגים כרכיב הנמדד).

תכליות אפשריות של הפרויקט הינם: כבסיס לגורמים במערכת החינוך לצורך קבלת החלטות בתקצוב והעברת משאבים, בהסתמך על הנתונים והתוצאות מהשנים האחרונות, מבט חיצוני על המערכת ותפקודה. כמו כן, כמנגנון הערכה של בתי ספר, ערים ונפות לאיתור הצטיינות, מגמות ופערים הדורשים מענה. בנוסף, ברמת בית הספר הדבר מייצר בסיס לשת"פ ולמידה הדדית החוצה מגזרים, עם בתי ספר מקבילים ברמת הקשיים או ההישגים.

ברקע, עבור הורים המתלבטים היכן להשתקע או תלמידים שמתלבטים לאיזה תיכון ללמוד, יכולים לאתר את העיר והתיכון המתאים להם בהתאם למאפיינים ספציפיים.

סיכום כללי -

בפרויקט לקחנו מאגרי מידע שונים והתנסנו במודלים וכלים שונים לניתוח נתונים, תוך התמודדות עם פערי מידע ונתונים חסרים, פערים בין מקורות שונים. הפרויקט היה מלמד מאוד על הכלים והמימושים שלהם ותוצאותיו משקפות את החשיבות בניתוח והעמקה בנתונים כשלב בסיס ובהצגה אפקטיבית ונוחה של המסקנות העולות מהן. קונקרטית, השימוש במודלים שונים לאפיון וחלוקה של בתי הספר ואיתור קורלציות ומאפיינים ברורים בסאב-סטים האלו, מוכיח את השימושיות של ניהול ומדיניות מוכווני נתונים ככלים שיאפשרו הבנה של מציאות מורכבת ודיוק של תהליכי עבודה.